THE ASTROPHYSICAL JOURNAL, 375:423–426, 1991 July 1

LINEAR AND CIRCULAR DIGITAL SPECTRAL ANALYSIS OF SERIAL DATA

RICHARD B. STOTHERS
Institute for Space Studies, NASA/Goddard Space Flight Center, 2880 Broadway, New York, NY 10025
Received 1990 July 11; accepted 1990 December 18

ABSTRACT

Two methods of digital spectral analysis of unevenly sampled data are developed and illustrated here. One method uses a linear function of time (or space), the other uses circular functions. The circular method turns out to be essentially equivalent to a least-squares sine-wave analysis. The linear, anharmonic method uses only the field of real numbers and elementary algebraic operations, and hence it can be made computationally very fast and accurate. Both methods are very general, properly handling all kinds of time series ranging from simple series consisting only of the times of events to complicated series consisting of pulses with long duty cycles. The two methods are here applied to the analysis of annual mean relative sunspot numbers.

Subject headings: numerical methods — Sun: sunspots

1. INTRODUCTION

Many experimental problems in astronomy and in other disciplines require spectral analysis of serial data sequenced in time or space. The methods of analysis previously used have essentially fallen into two broad categories: harmonic analysis (e.g., the Schuster periodogram, fast Fourier transform, least-squares sine-wave analysis, Blackman-Tukey power spectrum analysis, and maximum entropy spectral analysis) or folding methods. These techniques have been described well in numerous publications. Presented here are two new methods of analysis, one of which turns out to be essentially equivalent to least-squares sine-wave analysis while the other belongs in a category of its own. Both methods, however, are ultimately rooted in discussions of early physical problems and of early mathematical techniques devised for the solution of these problems. What is perhaps most important in this paper is the extension of these early techniques from the treatment of time series consisting only of a sequence of dates to the treatment of time series with amplitude information.

2. TIME SERIES WITHOUT AMPLITUDE INFORMATION: PREVIOUS WORK

To analyze his observations of simple pendulum motion, Galileo broke ground by measuring independently the amount of time required for one oscillation and the number of oscillations occurring in a fixed time (Fermi & Bernardini 1961). If N is the number of observed oscillations, his statistical model for this process can be represented by a periodic function that is linear in both the model parameters and the controllable variables,

$$t_i = \tau + n_i \Pi + e_i \,, \tag{1}$$

where τ is the initial phase, Π the period of oscillation, n_i an integer or zero, and e_i a small error term. In modern applications of equation (1), observed event times t_i ($i = 1, 2, \ldots, N$) are fitted to the model by using tabular, graphical, or analytical methods. If the n_i are known with respect to some arbitrary starting value and the e_i are normally distributed with mean 0 and variance σ^2, minimum-variance estimates of the parameters τ and Π can be derived by the method of least squares. Note that, if some event times are either clustered or missing, n_i will not in general be equal to i.

Often, however, the n_i are not observed quantities. Broadbent (1955, 1956) considered the case where τ is either known or assumed *a priori*; defining d_i (present notation) as the distance of t_i from the nearest predicted time $\hat{t}_i = t_0 + n_i P$, he discussed the statistical properties of the lumped variance of the observations,

$$s^2 = \left(\sum_{i=1}^{N} d_i^2 \right) \Big/ N \,, \quad d_i = t_i - \hat{t}_i \,, \tag{2}$$

as a function of the trial value P for the "quantum" Π, with $t_0 = \tau$. (An unbiased estimate of variance for $N - 1$ degrees of freedom, $s^2 N/(N-1)$, could alternatively have been adopted.)

Two rediscoveries and generalizations of Broadbent's method have allowed for the situation where τ is unknown. In these approaches, a best-fit value of τ is obtained for each P by adopting a set of trial initial phases $t_{0j} (\mathrm{mod}\ P) (j = 1, 2, \ldots, M)$, then calculating the d_{ij}, and finally locating the smallest member of the corresponding set of computed $s_j = (\sum_i d_{ij}^2/N)^{1/2}$ (Stothers 1979) or of computed $\delta_j = |\sum_i d_{ij}|/N$ (Raup & Sepkoski 1984). The latter approach, using δ_j, simply treats (δ_j, s_j^2) as a sufficient statistic, but my numerical experiments have shown that δ_j can attain exactly zero for more than one initial phase for some trial periods, even in the case of a perfectly periodic time series. Consequently, only the use of s_j yields a fully consistent, maximum-likelihood solution.

An entirely different approach is possible, as Schuster (1897) showed in his study of the statistical distribution of times of earthquakes and von Mises (1918) independently demonstrated in the case of the distribution of experimentally measured atomic weights. The measurements t_i are first mapped onto a circle by converting them to angles,

$$\theta_i = 2\pi t_i / P \ (\mathrm{mod}\ 2\pi) \tag{3}$$

(see also Kendall 1974). With the definitions

$$A = \left(\sum_{i=1}^{N} \cos \theta_i \right) \Big/ N \,, \quad B = \left(\sum_{i=1}^{N} \sin \theta_i \right) \Big/ N \,, \tag{4}$$

an application of circular statistics leads to a mean vector magnitude R (a normalized measure of goodness of fit) and a phase τ that minimize the dispersion at trial period P:

$$R = (A^2 + B^2)^{1/2} \,, \quad \tau = (P/2\pi) \tan^{-1} (B/A) \,. \tag{5}$$

The full range of R is 0 to 1. As in related harmonic-analysis

techniques, this approach requires calculating trigonometric functions, which is slow. More seriously, erroneous answers may be obtained by this method if the distribution of the s_j is asymmetric about its true minimum or has more than one minimum (as happens at some trial periods even in the case of perfectly periodic time series, although not at harmonics of the fundamental period).

3. INTERPRETATION OF THE SPECTRA

Calibration of the more accurate, and potentially faster, linear method (Stothers 1979) can be done by analyzing a finite Dirac comb, which is a periodic series of unit impulse functions. For simplicity, the $N - 1$ spacings between the N equispaced δ-function pulses are set equal to unity, and frequency $f = 1/P$ is used instead of period for the analysis. The nondimensional spectrum $s(f)f$ appears as a semi-infinite, horizontal continuum of height

$$s_c(f)f = [(N^2 - 1)/(12N^2)]^{1/2} , \tag{6}$$

broken by a sequence of "absorption dips" that descend to zero (§ 4). Below a critical frequency $f^* = 1/N$, $s(f)f$ monotonically decreases in proportion to f. To display the spectral transform of the time series as a sequence of "emission peaks" above a zero-valued continuum, a residuals index $S(f) = s_c(f)f - s(f)f$ is adopted, although an alternative choice might be a variance index, $V(f) = s_c(f)^2 f^2 - s(f)^2 f^2$. Normalized indices would be $S_n(f) = S(f)/s_c(f)f$ and $V_n(f) = V(f)/s_c(f)^2 f^2$.

The spectral peaks at frequencies $f > f^*$ are found by analytical or empirical means to obey the following "quantum rules":

1. Primary spectral peaks have the highest maxima, $S(f) = s_c(f)f$, and occur centered at the harmonic frequencies $f = 0, 1, 2, \ldots$, for $N \geq 2$. See § 4.

2. Secondary spectral peaks appear with maxima occurring at the subharmonic frequencies $f = 1/2, 3/2, 5/2, \ldots$, for $N \geq 3$. For small, even N, splitting of the peaks in two causes a shallow valley to appear at the subharmonic frequencies themselves. At large N, the maxima have $S(f) \approx 0.134\, s_c(f)f$. See § 4.

3. In analogous fashion, higher order series of spectral peaks arise successively for $N \geq 4$, but have extremely small $S(f)$ values at large N.

4. Competing with the development and positioning of the secondary and higher order series of spectral peaks is the occurrence of "line splitting" (subdivision) of the primary peaks for $N \geq 3$. The realizable number of subdivisions, however, is restricted by the total number of "allowed" spectral peaks between successive primary peaks, which number is $N - 2$ (only approximate for large N), and by the fact that subdivision invariably occurs in a symmetric pair about the primary peak. The members of the innermost pair always lie, with respect to the primary peak, at $\Delta f = \pm 3/(2N_0)$, where $N_0 = N$ (odd N) or $N_0 = N - \frac{1}{4}$ (even N). At large N, the innermost pair shows $S(f) \approx 0.134 s_c(f)f$, i.e., the same as for the subharmonic (secondary) spectral peaks. More distant satellite peaks have their maxima successively separated from their inner neighbors by $\Delta f \approx \pm 2/(2N_0)$, and exhibit amplitudes that decay with increasing distance from the primary peak. The phenomena of nonzero line width and of line splitting resemble slit diffraction patterns and are due to the finite length of the record. The diffraction analogy has been quantitatively worked out by Blackman & Tukey (1958) in the case of the Fourier transform of a finite Dirac comb.

5. Spectral peaks belonging to the secondary and higher order series of peaks also split, but obey quantitatively different separation laws for the components. The $S(f)$ values of the split components are very small at large N.

For a long series of observed event times that are not obviously periodic, the distribution of major and minor peaks in the spectrum can occasionally help to identify the spectral peak which refers most nearly to a fundamental period. Owing to the relative insensitivity of $s_c(f)f$ to N, the statistic $S(f)$ remains a well-behaved spectral index even for highly aperiodic series. However, $S(f)$ can be slightly negative at some trial periods for an aperiodic series.

4. EVALUATION AND MEANING OF THE SPECTRAL CONTINUUM $s_c(f)f$

Equation (6) represents the continuum of the spectrum $s(f)f$. The mathematical proof of equation (6) follows by the method of induction, starting with $N = 2$. It is convenient to evaluate the spectral continuum at $f = 1/N$, so that $[s_c(f)f]^2 = N^{-3} \min_j \{\sum_i d_{ij}^2; j = 1, 2, \ldots, M\}$. Simple inspection of the periodic series of times is sufficient to indicate for each low value of N the most favorable location for τ, such that $\sum_i d_{ij}^2$ will be smallest. Formation of the sums for successively higher values of N leads to the general result

$$[s_c(f)f]^2 = \begin{cases} 2N^{-3} \sum_{i=1}^{N/2} (i - \frac{1}{2})^2 , & N \text{ even} , \quad (7a) \\ 2N^{-3} \sum_{i=1}^{(N-1)/2} i^2 , & N \text{ odd} . \quad (7b) \end{cases}$$

Application of elementary integer arithmetic then yields equation (6). Note that $1/16 \leq [s_c(f)f]^2 \leq 1/12$ for $2 \leq N \leq \infty$.

In a spectrum plotted for the residuals index, $S(f) = s_c(f)f - s(f)f$, the quantity $s_c(f)f$ refers to the heights of all the spectral peaks at the harmonic (primary) frequencies. It is easy to show that, at subharmonic (secondary) frequencies, the heights of all the peaks approach $1/(12)^{1/2} - 1/(16)^{1/2} = 1/(12)^{1/2} \times 0.134\ldots$ as $N \to \infty$, and therefore, for large N, $S(f) \approx 0.134\, s_c(f)f$.

The quantity $s_c(f)f$ has considerably wider significance in that it also represents the expectation value of $s(f)f$ for the uniform probability distribution of an infinite number of random times. Previously, this distribution (alternatively called the rectangular distribution) was studied analytically and numerically by Broadbent (1955, 1956), but his analytical result, which is equivalent to $E\{s(f)f\} = s_c(f)f = 1/(12)^{1/2}$, refers strictly to $N = \infty$. My Monte Carlo simulations, using a sampling interval running from $t = 0$ to $t = N + 1$ such that the expected values of the N ordered random times are $1, 2, \ldots, N$, yield an expectation value of $s(f)f$ equal to $E\{s(f)f\} \approx s_c(f)f - (2N^{-1/2}) s_c(f)f$ for $f > f^*$.

This result is obviously relevant to hypothesis testing at *a priori* frequencies. For the null hypothesis of a Poisson process of completely random times, the expectation value of the residuals index is $E\{S(f)\} \approx (2N^{-1/2}) s_c(f)f$ for $f > f^*$.

5. TIME SERIES WITH AMPLITUDE INFORMATION

Time series that contain amplitude information $w_i(t_i)$ may be operationally regarded either as a sequence of unequally weighted times or as a sequence in which some of the times

coincide. Either of these interpretations permits an extension of the preceding methods, because they validate adopting the statistic $S(f) = s_c(f)f - s(f)f$ with

$$s(f)^2 = \min_j \left\{ \sum_{i=1}^{N} w_i d_{ij}^2 \Big/ \sum_{i=1}^{N} w_i \; ; \quad j = 1, 2, \ldots, M \right\} \quad (8)$$

or the analogous circular statistic $R(f) = (A^2 + B^2)^{1/2}$ with

$$A = \sum_{i=1}^{N} w_i \cos \theta_i \Big/ \sum_{i=1}^{N} w_i , \quad B = \sum_{i=1}^{N} w_i \sin \theta_i \Big/ \sum_{i=1}^{N} w_i . \quad (9)$$

In the latter case, $R(f)$ apart from a different normalization constant is equivalent to the spectral amplitude in least-squares sine-wave analysis in its lowest order approximation (e.g., Barning 1963), which itself is essentially equivalent to the spectral amplitude in a slightly modified version of ordinary periodogram analysis (e.g., Lomb 1976; Scargle 1982). The original periodogram method was devised only for evenly sampled data (Schuster 1898). The present derivation, however, is new and illustrates why the use of $R(f)$ is as appropriate for a series of very narrow pulses as for a series of sinusoidal pulses.

Note that every observation enters the analysis with its exact time and with its full weight. No interpolation is necessary to fit a mesh of equally spaced times. Furthermore, the absence of a need to bin the observations means that any associated losses of resolution and of information are not incurred.

To calibrate the new method of *weighted linear* spectral analysis, a finite series of periodic pulses with a long duty cycle will be used, specifically, a pure cosine curve, $w = 1 + \cos(2\pi t)$. Over the practical frequency range $f > f^*$, the spectral transform $S(f)$ shows only one supremum: a high, wide-band peak centered at the fundamental frequency $f = 1$, which is surrounded by side lobes of diminishing amplitude, representing the spectral leakage. As the number of cosine cycles increases, the side lobe density also increases and the main spectral peak becomes narrower. The frequencies of the maxima of the innermost side lobes are very close to those arising from a finite Dirac comb if N is equal to the number of cosine cycles plus one-half. The case of eight equispaced pulses is illustrated in Figure 1 for both the pure cosine curve and the finite Dirac comb.

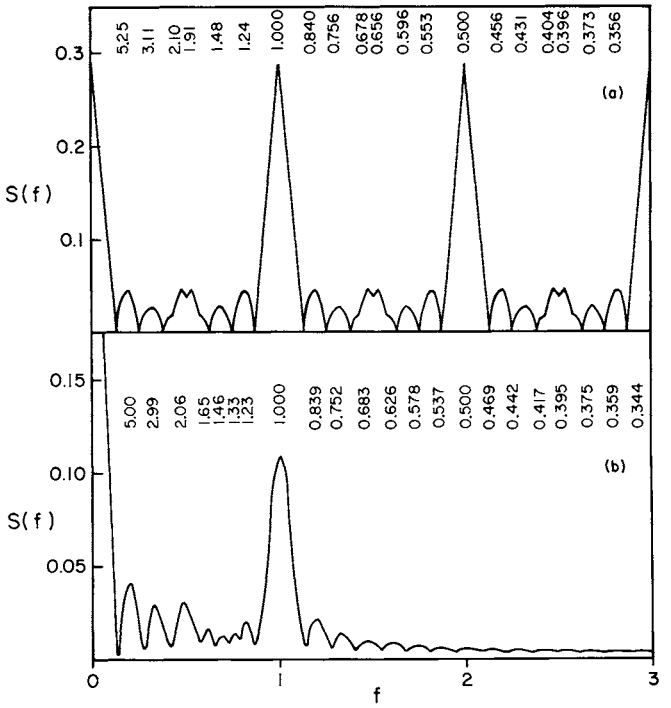

FIG. 1.—Spectra $S(f)$ for (*a*) a finite Dirac comb consisting of eight pulses and (*b*) a pure cosine curve containing eight maxima (seven and a half cycles). Each cycle in the cosine curve has been sampled at 64 equal time intervals. A higher sampling rate leads to little change. Periods of the major peak maxima are listed in units of the fundamental period, II. Results based on $R(f)$ are rather similar to those for $S(f)$.

6. COMPUTATION TIME

The computation time on a scalar processing machine scales as $NN_f N_\phi$ for the linear method and as NN_f for the circular method. Here N is the number of observations, N_f is the number of trial frequencies, and N_ϕ (called M above) is the number of trial phases per trial frequency. In both methods, N, N_f, and N_ϕ are mutually independent.

On a scalar machine, the need to circulate through N observations for every trial phase makes the linear method computationally slower than the circular method in most problems for which a reasonable number of trial phases (and trial frequencies) is required. For example, using a program written in FORTRAN for the Goddard Institute for Space Studies' Amdahl 5870 sequential processor with the MVS/XA operating system, linear spectral analysis of 26 observations with 6000 trial frequencies and with 20 trial phases per trial frequency takes ~ 4 times longer than does the same computation using circular spectral analysis.

On the other hand, a parallel or vector processor can calculate simultaneously all the needed trial phases. With program vectorization, the linear method clearly becomes faster than the circular method. Although clever ways to speed up the circular method (which is essentially equivalent to the modified periodogram method) are now available (e.g., Kurtz 1985; Press & Rybicki 1989), the need to calculate or look up sine and cosine functions may ultimately keep this method slower than the vectorized linear method, which seems to be close to the ultimate in digital simplicity. The new algorithm could even be easily programmed in machine language.

7. APPLICATIONS

Some aspects of the power of the linear method can be demonstrated with a classical time series: the sequence of annual mean relative sunspot numbers since the year 1700 (Waldmeier 1961; US Department of Commerce 1987). Transform spectra are displayed in Figure 2 for three comparative cases: (*a*) linear spectral analysis of the 26 dates of sunspot maximum between 1700 and 1986; (*b*) weighted linear spectral analysis of the whole sunspot curve, 1700–1986; and (*c*) weighted circular spectral analysis of the same curve.

The results for the periods of the spectral peaks compare very favorably in this complicated test case, in spite of the enormous differences inherent in the three spectral methods used. The main spectral peak occurs at a period of ~ 11 yr, and some of the minor peaks exhibit amplitudes and periods that indicate that they are not side lobes of the main peak. Berger, Mélice, & van der Mersch (1990) recently used on the same time series six different spectral methods: classical harmonic analysis, Blackman-Tukey power spectrum analysis, periodic regression, maximum entropy analysis, minimum cross-entropy analysis, and Thomson multitaper analysis. Not surprisingly, at the higher harmonics of the record length where

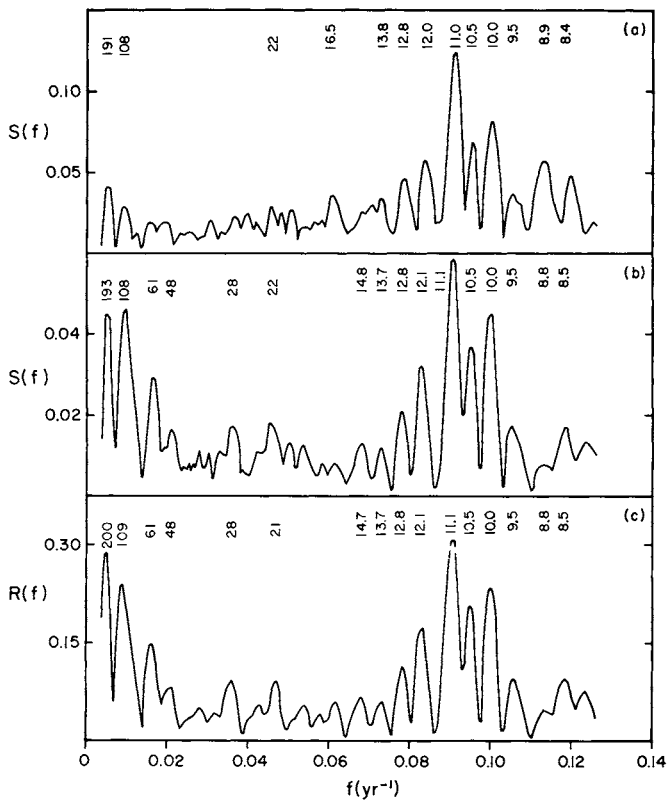

FIG. 2.—Transform spectra of annual mean relative sunspot numbers, 1700–1986, based on: (*a*) linear spectral analysis of the 26 dates of sunspot maximum, (*b*) weighted linear spectral analysis of the whole sunspot curve, and (*c*) weighted circular spectral analysis of the same curve. The raw data and the computed spectra have not been detrended or filtered. Periods of the major peak maxima are listed in units of years.

the frequencies are very closely spaced, the results based on classical harmonic analysis are virtually identical to the present results based on circular spectral analysis. An equivalent degree of resolution seems to be achieved in the cases of periodic regression and Thomson multitaper analysis. Although it is not the purpose of this paper to interpret these numerous spectral peaks, it cannot escape notice that many of the minor peaks must be relatively independent of the actual details of the sunspot curve, since they occur even when only the dates of sunspot maximum are analyzed (Fig. 2*a*).

8. CONCLUSION

Since the linear and circular methods of analysis are not restricted to equally spaced dates, they are readily able to handle gaps in data acquisition. As long as enough cycles are available and nonstationarities are not dominant, the central limit theorem ensures that it is possible to locate approximately the correct mean period; consequently, the distributions of the errors e_i need not be either homoscedastic or normal (which in general they cannot be owing to the physical constraint $|d| \leq \frac{1}{2} \max \{t_{i+1} - t_i\}$). The potentially great computing speed of the linear method in a parallel or vector processor may commend it over analog methods for some kinds of real-time data processing and also for performing nonparametric tests of significance that employ Monte Carlo simulations. The crucial difference between the two basic types of spectral methods is that linear analysis fits a single, linear, discrete function for time, whereas circular and other forms of harmonic spectral analysis fit, in effect, a multiple, sinusoidal, continuous function for amplitude.

This paper is dedicated to the memory of Sergej A. Lebedeff, who helped to spark in me an interest in time series analysis. An anonymous referee made some useful suggestions for greater clarity in various parts of the paper. The listing of a FORTRAN program to calculate $S(f)$ and $R(f)$ is available upon written request to the author (no requests for disk, tape, or electronic transfer, however, will be honored). The program code is not vectorized for use in parallel or vector processors.

REFERENCES

Barning, F. J. M. 1963, Bull. Astr. Inst. Netherlands, 17, 22
Berger, A., Mélice, J. L., & van der Mersch, I. 1990, Phil. Trans. Roy. Soc. London, A330, 529
Blackman, R. B., & Tukey, J. W. 1958, The Measurement of Power Spectra from the Point of View of Communications Engineering (New York: Dover)
Broadbent, S. R. 1955, Biometrika, 42, 45
———. 1956, Biometrika, 43, 32
Fermi, L., & Bernardini, G. 1961, Galileo and the Scientific Revolution (New York: Basic Books)
Kendall, D. G. 1974, Phil. Trans. Roy. Soc. London, A276, 231
Kurtz, D. W. 1985, MNRAS, 213, 773
Lomb, N. R. 1976, Ap&SS, 39, 447
Press, W. H., & Rybicki, G. B. 1989, ApJ, 338, 277
Raup, D. M., & Sepkoski, J. J., Jr. 1984, Proc. Nat. Acad. Sci., 81, 801
Scargle, J. D. 1982, ApJ, 263, 835
Schuster, A. 1897, Proc. Roy. Soc. London, 61, 455
———. 1898, Terrestrial Magnetism, 3, 13
Stothers, R. 1979, A&A, 77, 121
US Dept. of Commerce 1987, Sol. Geophys. Data (Part 1), 515, 11
von Mises, R. 1918, Phys. Zs., 19, 490
Waldmeier, M. 1961, The Sunspot-Activity in the Years 1610–1960 (Zurich: Schulthess)